DOCUMENT RESUME

ED 423 714 FL 801 252

AUTHOR Kahn, Andrea B.; Butler, Frances A.; Weigle, Sara Cushing; Sato, Edynn

TITLE Adult English-as-a-Second-Language Assessment Project. Final Report: Year 3.

SPONS AGENCY California State Dept. of Education, Sacramento.
PUB DATE 1995-12-00
NOTE 46p.; For related documents, see FL 801 250-253.
CONTRACT 3151
PUB TYPE Reports - Descriptive (141)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adult Education; *English (Second Language); Language Proficiency; Literacy Education; Program Descriptions; Second Language Instruction; State Standards; Statewide Planning; *Student Placement; Testing
IDENTIFIERS California; *Placement Tests

ABSTRACT

The report describes activities in the final year of a 3-year project designed to address the placement needs of adult education agencies teaching English-as-a-Second-Language (ESL) students in California. The long-term goal was to identify instruments appropriate for use with the state's model standards for language proficiency. The two primary tasks for the project's third year were: (1) establishment of initial cutoff ranges for the commercially available instruments recommended for use based on the first two years' work, and (2) creation of a test development plan to guide production of operational instruments for placing students according to state standards. Appended materials include: a list of working group members; general content grids for reading, writing, and listening tests; a sample questionnaire used to survey teachers and students during the field testing phase; and data from a special pilot administration of reading, writing, and listening test items. Contains 9 references. (MSE) (Adjunct ERIC Clearinghouse on Literacy Education)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 714

CSE

Center for the Study of Evaluation

California Department of Education

Adult English-as-a-Second-Language Assessment Project

Final Report: Year 3

Andrea B. Kahn, Frances A. Butler,
Sara Cushing Weigle, and Edynn Sato

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Jr Richard Stiles

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Center for the
Study of Evaluation

UCLA Graduate
School of Education

405 Hilgard Avenue
Los Angeles, CA 90024-1522
(310) 206-1532

FL 801252

2

BEST COPY AVAILABLE

California Department of Education

Adult English-as-a-Second-Language Assessment Project

Final Report: Year 3

Andrea B. Kahn, Frances A. Butler,
Sara Cushing Weigle, and Edynn Sato

Project Director: Frances A. Butler

December 1995

Center for the Study of Evaluation
Graduate School of Education & Information Studies
University of California, Los Angeles

The work reported in this document was conducted by the UCLA Center for the Study of Evaluation under Contract No. 3151, a state-administered contract of the ADULT EDUCATION ACT, P.L. 100-297 as amended, Section 353, from the California Department of Education, 721 Capitol Mall, Sacramento, CA 95814. However, the opinions expressed do not necessarily reflect the position or policy of that department or the U.S. Department of Education. No official endorsement of this work should be inferred.

Acknowledgments

The work carried out during the third and final year of this project was accomplished because many skilled professionals were willing to assist at various stages in the development of prototype items. The number of people who contributed to the prototype development effort highlights the complexity of test development processes in general. Project staff could not have brought this undertaking to fruition without the assistance of the people mentioned below. To all we extend our sincere gratitude.

Dr. Raymond Eberhard, Administrator, Adult Education Policy and Planning Unit, California Department of Education (CDE), provided us with the opportunity to create and articulate a systematic approach for linking standards and assessments in an adult English-as-a-Second-Language (ESL) environment. He demonstrated his confidence and support by giving us free rein to use our expertise in producing an assessment model for California that could have impact in other adult ESL venues in the United States.

The ESL Assessment Working Group members were an integral part of the process and provided feedback and guidance at every stage from conceptualization to actually piloting the items. As we have said before, they were our partners in the truest sense of the word. This project could not have happened without them.

Lynn Drew Bartlett, consultant in the Adult Education Policy and Planning Unit, CDE, served as the project monitor during this third year of work. She met periodically with project staff and gave thoughtful input to planning efforts for each stage of work. In addition, she reviewed drafts of all project documents and provided meticulous feedback.

Dr. Jean Turner of the Monterey Institute of International Studies, a language testing expert, served as a consultant to the project during the second and third years. Her insight and feedback were invaluable in helping us frame our discussions and revise items and specifications in ways that we believe are relevant to both adult education agencies and future test developers.

Eighteen agencies allowed us to look at their current placement tests and thereby helped to provide insight into testing material for adult education students.

Jayme Adelson-Goldstein, Melisa Haines, Sue Larsen, Ronna Magy, Patricia Snyder, Cynthia Taskesen, and Chris Waldeck helped to score the pilot testing writing samples. All participated in training and scoring sessions at UCLA and provided feedback on the effectiveness of the scoring rubric as part of the process.

The following assisted with the development of listening prototypes by writing items or assisting in the production of accompanying audiotapes: Bob Agajeenian, Earl Davis, Ron Dietel, Todd Hanson, Amy Lee, Lyn Repath-Martos, John Novak, Joe Plummer, Laura Singer, Eileen Tebbs, Gilbert Winstead, Clarke Weigle, and Patric Z.

Kathleen Ford provided guidance in recording and editing the audiotapes. The combination of her expert knowledge and her willingness to work with us made her a pivotal participant in the development of the listening items.

Andrew M. Park produced the drawings that accompany the listening items. Andy was able to listen to our ideas, look at our stick figure sketches, and return with polished drawings that were just what we had in mind. We are pleased that we could avail ourselves of his talent.

Lynne Nash provided invaluable assistance by coordinating the production of the audiotapes and by overseeing scoring and data entry for all of the pilot testing. She carried out these tasks in a seemingly effortless manner and, by doing so, freed us to pursue the next steps in the development process.

Jamie Schubiner produced the graphics and formatted the exercise booklets. Howard Kim and Kim-An Nguyen aided in scoring the exercises and with data entry. Thuy-Van Le provided administrative support and assisted in the final formatting of this document.

Finally, the cooperation of all of the teachers and students at the thirteen agencies represented by the working group was vital to the prototype development process. They readily agreed to assist us in the development of the prototype items, the students by taking the exercises and providing feedback on the items; the teachers by providing input regarding item content and format.

Table of Contents

Introduction 1

Test Development Plan 3

Placement Process for Model Standards Levels 3

Specifications for Reading, Writing, and Listening 6

Abstracting and systematizing Model Standards content 6

Prototype development 9

Guidelines for Test Development 17

Conclusion 20

References 21

Appendix A: Adult ESL Assessment Working Group Members 23

Appendix B: General Content Grids 27

Appendix C: Sample Teacher/Student Questionnaire 35

Appendix D: Special Pilot Administration of Reading, Writing, and Listening Items 39

List of Figures and Tables

Figure 1: ESL Placement Process 4

Figure 2: Specification Development Process 7

Table 1: Summary of reading/writing contexts pilot tested in October 1994 by test level 12

Table 2: Number of students in the pilot administration of reading/writing contexts by test level and Model Standards proficiency level 12

Table 3: Summary of listening contexts pilot tested in March 1995 by test level 15

Table 4: Number of students in the pilot administration of listening contexts by test level and Model Standards proficiency level 15

Table D1: Number of items in May 1995 pilot administration by test level and skill area 41

Table D2: Number of students in May 1995 pilot administration of reading, writing, and listening items by test level and Model Standards proficiency level 41

Table D3: Descriptive statistics for May 1995 pilot administration of Level A items by Model Standards proficiency level and skill area 42

Table D4: Descriptive statistics for May 1995 pilot administration of Level B items by Model Standards proficiency level and skill area 42

Table D5: Correlations among skill areas for May 1995 pilot administration of Level A reading, writing, and listening items 43

Table D6: Correlations among skill areas for May 1995 pilot administration of Level B reading, writing, and listening items 43

Introduction

This document reports on work conducted by the UCLA Center for the Study of Evaluation during the third and final year of the Adult English-as-a-Second-Language (ESL) Assessment Project supported by the California Department of Education (CDE). The project was designed to address the placement testing needs of adult education agencies[1] in the state that were in the process of implementing the California *English-as-a-Second-Language Model Standards for Adult Education Programs*[2] (California Department of Education, 1992). The impetus for this project grew out of the need to facilitate agency implementation of the Model Standards through the use of placement instruments that match the standards in both content and approach. Since the Model Standards emphasize integration of skill areas, communicative language learning, and the use of multiple measures for assessment purposes, instruments used to place students into appropriate levels must have the same orientation. Thus, the long-term goal of the project was to identify a variety of instruments appropriate for use with the Model Standards, thereby providing a menu of tests from which agencies could select to satisfy individual needs.

Critical to the success of the project was the partnership established at the onset between project staff and the ESL Assessment Working Group[3] consisting of representatives from 13 agencies across the state. (See Appendix A for the list of working group members during Year 3.) The interaction at every juncture between project staff and the working group

---

[1]Henceforth in this document, adult education agency or agencies in California will be referred to as "agency" or "agencies."

[2]Henceforth in this document, the *English-as-a-Second-Language Model Standards for Adult Education Programs* will be referred to as the Model Standards. There are seven proficiency levels designated in the Model Standards: beginning literacy, beginning low (BL), beginning high (BH), intermediate low (IL), intermediate high (IH), advanced low (AL), and advanced high (AH). The Adult ESL Assessment Project addresses placement only into levels beginning low through advanced high.

[3]Henceforth in this document, the ESL Assessment Working Group will be referred to as the working group. The working group was supported by the CDE and played an active, vital role in the project work. (See Butler et al., 1993, pp. 3-4, for a detailed explanation of the role of the working group.)

members helped to assure the quality and appropriateness of the work for the adult ESL population.

The first year of work involved the review of 18 commercially available instruments to determine their suitability in terms of content match with the Model Standards. From the 18 reviewed, five potentially promising instruments were identified and field tested to determine the range of each instrument vis-à-vis the Model Standards proficiency levels and to reassess the content in light of student performance on the items. (See Butler, Weigle, & Sato, 1993, for a detailed report of Year 1 work.)

The second year of work included analysis and interpretation of the field testing results from Year 1, a survey of agencies across the state to document current ESL placement practices, and the development of a framework for producing assessment models, typically referred to as prototypes. Weigle, Kahn, Butler, and Sato (1994) provide in-depth analysis of the field testing results, discuss a recommended placement process to provide a context for the prototypes, and include guidelines for prototype development. Kahn, Butler, Weigle, and Sato (1994) provide the results of the survey on placement procedures in California.

There were two primary tasks for the third year of work. The first involved establishing initial cutoff ranges for the commercially available instruments that, on the basis of the field testing results, were recommended for use with the Model Standards; the second involved the creation of a test development plan to guide the production of operational instruments for placing students into the levels defined by the Model Standards.

Two of the five instruments field tested in Year 1 were recommended for inclusion on the proposed menu: the Basic English Skills Test (BEST) and the New York State Place Test (NYS Place Test). Both instruments were re-administered at agencies across the state to gather data for recommending initial cutoff ranges. The BEST was field tested with 180 students from beginning low through intermediate high at three different agencies.[4] The NYS Place Test was field tested with 243 students from beginning low through advanced high at four different agencies. The

[4]The content review and initial field testing results suggested that the BEST was only appropriate for the Model Standards proficiency levels beginning low through intermediate high. See Weigle et al., 1994, for a detailed discussion.

administration procedures and field testing results are reported in Weigle (1995).

The purpose of a menu of tests compatible with the Model Standards is to provide options across all four skill areas. While the BEST and the NYS Place Test provide viable options for assessing speaking ability, none of the commercially available instruments reviewed provide adequate coverage in their current form for assessing reading, writing, and listening (Weigle et al., 1994). A test development plan was therefore created to address the need for additional placement instruments that tap these skill areas. The test development plan (Butler, Weigle, Kahn, & Sato, 1996) incorporates information from the first two years of project work specifically with regard to the field testing results and the agency needs which emerged from the survey of current placement procedures. The plan contains specifications for developing reading, writing, and listening items as well as general guidelines for item and whole-test development and is the focus of the remainder of this document.

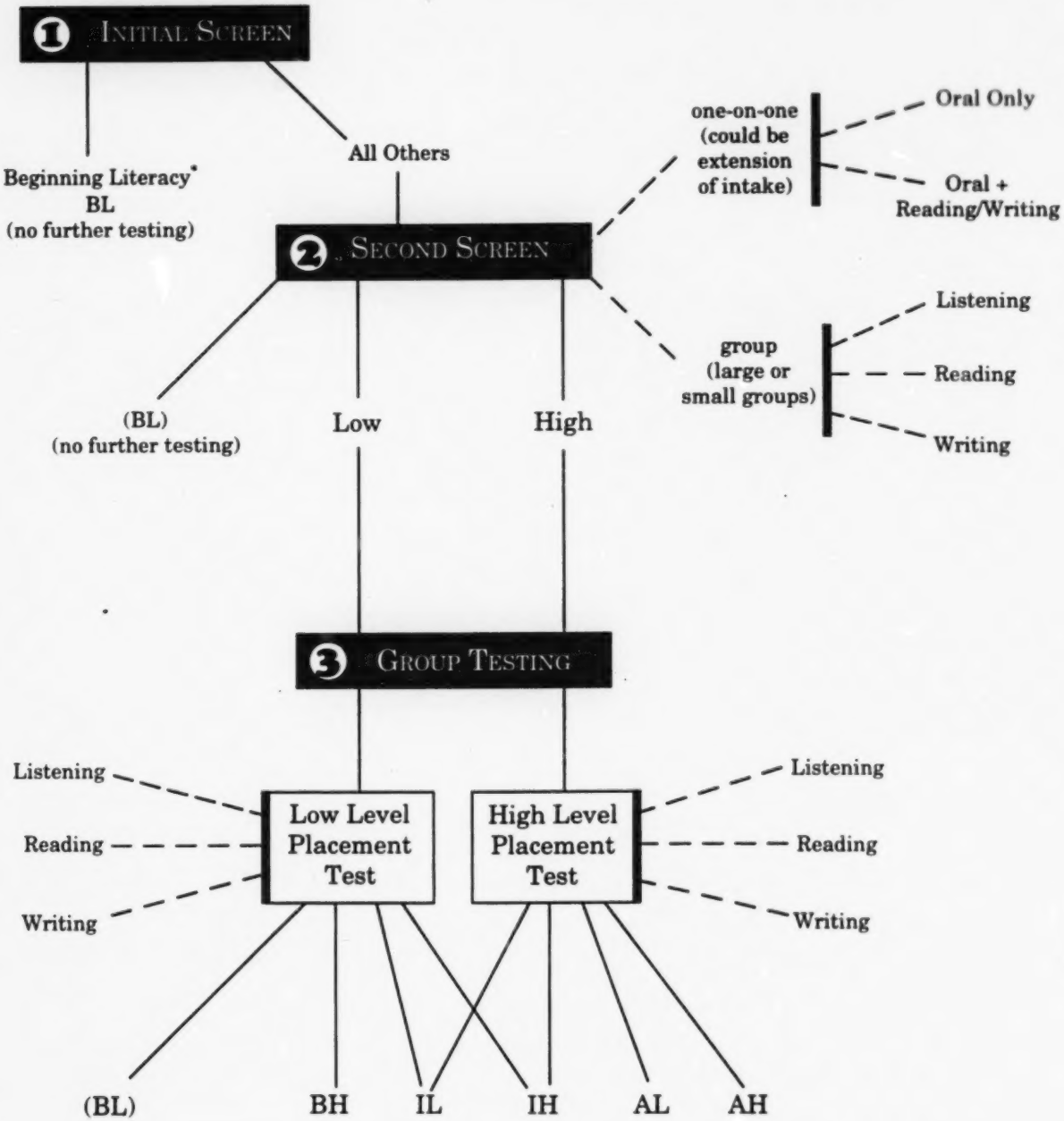
Test Development Plan

This section reports on the key components of the test development plan and is organized in the following way. First, the placement process is described to provide context for the test development plan. The placement process is followed by a discussion of the work that led to the development of text and item specifications for reading, writing, and listening. Finally, the general guidelines presented in the test development plan are summarized.

Placement Process for Model Standards Levels

The placement process was drafted during the second year of the project in order to develop specifications that match both the content of the Model Standards and the needs of adult education agencies. Two overriding issues led to the development of the placement process: 1) the need for group-administered tests that can be easily scored and 2) the difficulty of attempting placement into six levels with only one instrument. The three-tiered process, presented schematically in Figure 1, addresses both issues and provides the framework for test development. Key features of the process are discussed below.

Figure 1
ESL Placement Process



*There are seven proficiency levels designated in the Model Standards: beginning literacy, beginning low (BL), beginning high (BH), intermediate low (IL), intermediate high (IH), advanced low (AL), and advanced high (AH).

An initial screen identifies beginning literacy and beginning low students.

Most agencies already conduct some form of intake interview for administrative purposes and often use this process to identify those students with minimal or no literacy or oral skills. These students are usually placed immediately into beginning literacy or beginning low and no further testing is required. Inclusion of the intake interview as an initial screen serves to formalize its function in the overall placement process.

A second screen directs students to either low- or high-level testing.

The second screen is intended to be a quick procedure to make gross distinctions between lower and higher proficiency students. It will identify additional beginning low students, who will not be required to undergo further testing, and will direct all other students to appropriate low- or high-level placement tests. To accommodate varying agency needs, the second screen will be agency specific in terms of format and skill area focus. Some agencies may decide to use a group-administered test which could involve listening, reading, or writing, while others may prefer to make the second screen an extension of the intake interview by including a few oral questions, a short reading passage, or a simple writing task.

Final placement into levels is based on low- and high-level instruments that can be group administered.

Since most beginning low students will be identified by the first or second screen, low-level instruments will be used primarily to place students into beginning high, intermediate low, and intermediate high. High-level instruments will place students into intermediate low, intermediate high, advanced low, and advanced high. An important feature of this process is that both low- and high-level instruments allow for placement into the intermediate levels should the second screen fail to direct a student to the most appropriate level test.

The placement process described above is suggested as a model to help agencies place students into appropriate proficiency levels. In addition, it provided a framework for generating specifications and guidelines for test development.

Specifications for Reading, Writing, and Listening

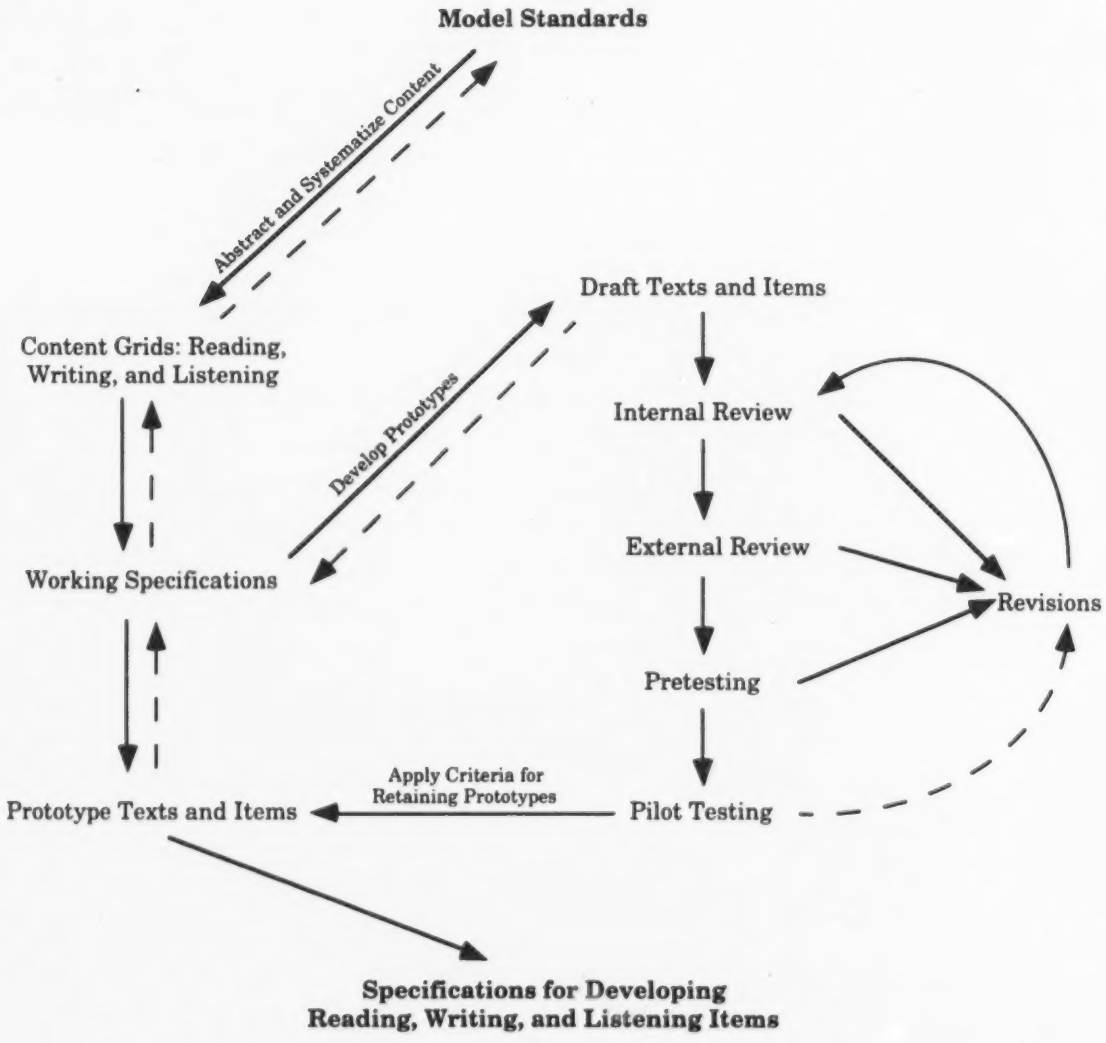
The specifications presented in the test development plan are intended to guide item writers in producing reading, writing, and listening items appropriate for the final stage of testing in the placement process described above. They were developed systematically through a process which began with careful analysis of the Model Standards, the content base for the test development effort.

Figure 2 is a graphic representation of the move from the Model Standards to the specifications and highlights the iterative nature of the development process. The first stage entailed content (i.e., topic areas, text types, and skills) being abstracted from the Model Standards and synthesized in the form of content grids. The content grids then became the basis for the working specifications which, along with the prototyping guidelines discussed in Weigle et al. (1994), guided the development of prototype texts and items. Generating prototypes involved texts and items being drafted, reviewed, revised, pretested, revised again, and pilot tested on a large scale. Only those texts and items that met established criteria were retained as prototypes, though information gleaned from the development of all texts and items was incorporated into the final specifications. A key component of the specification development process was that each stage could be revisited as new information was gathered; this was critical because implementation often shed light on unanticipated problems or constraints and ultimately allowed for greater precision in the specifications.

The description above provides an overview of the specification development process. What follows is a discussion of the two key stages in that process. The first focused on abstracting and systematizing the Model Standards content; the second involved the selection and development of prototype texts and items. Each is discussed in turn below.

Abstracting and systematizing Model Standards content. In order to determine test content for reading, writing, and listening, text types and skills were abstracted from the Model Standards and systematized in the form of a general content grid for each skill area. The general content grids, presented in Appendix B, contain information about

Figure 2
Specification Development Process



the type of texts and skills that are relevant for testing at each of the six Model Standards proficiency levels and serve as the content base for any type of assessment with the Model Standards—placement, diagnostic, progress, or exit. In addition, the general content grids contain information about language forms and functions that are relevant at each proficiency level.

To ensure that the text types and skills as categorized in the general content grids were an accurate reflection of the content of the Model Standards, a subcommittee of the working group met in July and August of 1995 to begin a validation process. Subcommittee members were asked to verify the categorization of text types and skills by systematically matching them to the Model Standards skill area descriptors. Once the text types and skills were verified, members were asked to determine whether the levels indicated for each text characteristic or skill were appropriate either because (a) the characteristic or skill was explicitly mentioned in the Model Standards for a given level or (b) the members felt that the characteristic or skill was relevant at that level based on their experience with adult school programs in the process of aligning to the Model Standards. Where there was lack of agreement, modifications were made to the text types and skills to reflect more accurately the Model Standards content. Modifications underwent further review until consensus was reached.

The next step in the validation process was to identify the text types and skills appropriate for placement from the general content grid. To do this, subcommittee members were asked to prioritize skills for placement by determining whether a given skill was essential, optional, or inappropriate for low and high level instruments, designated as Level A and Level B respectively. Decisions were based on the following primary considerations: (1) whether the skill is essential to determining a test taker's ability in a given skill area and (2) whether the skill can feasibly be assessed on a placement test given limited testing time and other operational constraints. Disagreements among members were discussed until consensus was reached.

Although this validation process was initiated for all three skill areas—reading, writing, and listening—it was only completed for reading and writing. This was due in part to the reading and writing skills being more fully articulated than the listening skills in the Model Standards.

With reading and writing, abstracting content from the Model Standards was direct and clear, which facilitated the categorization of skills and the definition of underlying constructs. With listening, abstracting content was more complicated. While the various listening settings that adult ESL students need to function in are clearly delineated in the Model Standards, the listening skills lack the same degree of specificity, which contributed to the difficulty of defining underlying listening constructs and verifying their match to the Model Standards. Models of listening performance found in the research literature and insight gained from small-scale tryouts of various item types informed the development of a schema for categorizing the listening skills. In addition, working group members reviewed the categories and definitions and made suggestions which were incorporated into the final specifications. The information gained through the process of more fully specifying the listening skills for test development purposes could inform future revisions of the Model Standards and thereby help reinforce the link between the tests produced from the specifications and the Model Standards proficiency levels.

Prototype development. The first step in the prototype development process involved the identification of appropriate topic areas and text types for placement testing. At a meeting in June 1994, working group members helped identify potential topics and source materials from those mentioned in the Model Standards and made suggestions for additional sources that might be suitable. Topics were selected from general content areas familiar to adult ESL students given their goals and experiences. These topics included, but were not limited to, shopping, banking, housing, health, transportation, current events, and community resources. Some common vocational topics such as employment and general workplace safety were also considered appropriate for placement testing.

Once possible topics were identified, texts were selected for prototype development and adapted as necessary. Although all texts were selected from materials originally prepared for a general English-speaking audience, some modifications were necessary for testing purposes. For example, proper nouns were changed to avoid association with actual people or organizations, visuals were added or modified to help orient the test taker to the text, and some texts were edited for simplicity or clarity, or

to make a given item type possible. In this case, an attempt was made to ensure that (1) the text remained as close to its original format as possible and (2) connected discourse retained a natural flow. Adapted texts were then reviewed by working group members to evaluate their appropriateness for use at specific proficiency levels given factors such as familiarity of topic, visual aids, and complexity of content, vocabulary, or syntax. Revisions were made when necessary and resubmitted for approval.

The next step in the process was to draft item types for each text. Survey results and discussions with the working group emphasized the need for item types that require the test taker to do more than select the correct answer and at the same time can be scored easily. In June 1994, working group members provided feedback on potential item types in terms of their appropriateness for adult ESL students and feasibility of scoring. Those identified as promising were explored and items were drafted for each selected text with two primary considerations in mind: (1) to tap a range of skills and proficiency levels as specified in the content grids, and (2) to try out a variety of formats to determine those most effective for assessing specific skills. Each context (i.e., text and accompanying items) was then reviewed internally to assure text appropriateness and quality of items.[5]

After making revisions prompted by the review, each context was pretested to determine whether the items and directions as formulated were comprehensible for the test taker. Pretesting provided an initial indication of the amount of time needed for students to work through a text and items and helped identify potential problems such as wording, familiarity with response formats, and task clarity. It also provided critical information regarding the feasibility of scoring a variety of item formats. With writing tasks, pretesting showed whether a given prompt would elicit a ratable sample.

Each context was pretested at one to three agencies, ordinarily with one class per level at each agency. Six contexts on which both reading and writing items were based were pretested from June through September

---

[5]For this project, internal review refers to project staff and working group members who were involved with the project on an on-going basis and external review refers to the outside language testing expert and teachers from the thirteen representative agencies who were not involved in the development process.

1994. Eight listening contexts were pretested, two in July 1994 and the remainder from October 1994 to February 1995. Only one or two contexts were pretested in a given class period so that time was not a factor in student ability to respond to the questions. Moreover, limiting the number of contexts pretested at one time allowed project staff to obtain specific feedback from teachers and students regarding the content and format of the texts and items.

The information gained from pretesting informed revisions and helped determine those contexts that should be retained and those that should not. For example, one of the reading/writing contexts was considered to be inappropriate due to the nature of the content and was therefore eliminated. Another underwent extensive revision for the opposite reason; the text lacked clarity and needed to be more thoroughly developed, but contained relevant content. This context was initially very difficult for students, even those at the advanced levels. However, student reaction to the text, a newspaper article about an adult ESL student much like themselves, was extremely positive. Students also felt the related writing task reflected the type of writing they need to be able to do. Therefore, in spite of the initial poor performance of the reading items pretested, the context was retained and several revisions were made to improve text and item clarity.

The goal of pretesting was to collect information at the task and item level that would inform revisions and help produce sets of items that were as strong as possible for pilot testing. While some contexts, such as the reading/writing context described above, were pretested and revised several times before a satisfactory set of items was obtained, other contexts required only a single pretest administration. Once pretest data were analyzed and revisions completed, texts and items were reviewed externally by a language testing expert in preparation for the pilot testing effort.

Pilot testing involved the administration of texts and associated items to a large number of students across an appropriate range of levels so that, in addition to content issues, statistical analyses could be run to determine whether the items were performing as expected for placement purposes.

Pilot testing of reading/writing contexts took place in October 1994.[6] Of the six contexts pretested, five were retained for pilot testing. Table 1 presents the contexts pilot tested at Level A (BL-IH) and Level B (IL-AH) and indicates the number of reading and writing tasks associated with each context. Level A reading/writing contexts were pilot tested with 570 students from beginning low through intermediate high at six different agencies. Level B reading/writing contexts were pilot tested with 658 students from intermediate low through advanced high at seven agencies. Table 2 provides the number of students in the pilot administration of Level A and Level B reading/writing contexts by proficiency level.

Table 1
Summary of reading/writing contexts pilot tested in October 1994 by test level

Level	Reading/Writing Context	Number of Associated Reading Items	Number of Associated Writing Tasks
A	Public Announcement	4	1
A	Bicycle Advertisements	10	0
A & B	Short Newspaper Article	4	1
B	Apartment Guide	10	0
B	Long Newspaper Article	9	2†

†Although two writing tasks were associated with this context, most students were asked to respond to one or the other. A small number of students responded to both tasks for comparison purposes. (See Kahn, forthcoming, for results.)

Table 2
Number of students in the pilot administration of reading/writing contexts by test level and Model Standards proficiency level

	Model Standards Proficiency Level							
Test Level	BL	BH	IL	IH	AL	AH	Visa†	Total
A	110	147	170	143				570
B			152	151	233	109	13	658

†Visa students resemble exit-level students (those more proficient than AH) and were administered the exercises to gauge the appropriateness of the tasks for the population.

[6]Reading and writing tasks were developed around the same source material to address the integration of skills emphasized in the Model Standards and were thus pilot tested at the same time. Listening items were developed and pilot tested separately.

For each reading/writing context, a variety of items using both constructed and selected response formats[7] were developed to assess the following reading skills: locate specific information, draw meaning, extract and combine information, interpret relationships, and make inferences. Pilot testing results indicated that the items developed generally did a good job of discriminating across levels, particularly at the beginning and advanced levels. Item performance was somewhat less predictable at the intermediate levels, which may be a function of the items or the alignment process at the agencies tested.

Several of the item formats pilot tested proved to be promising and were thus included in the reading item specifications. For selected response, promising item formats include sequencing activities which require the test taker to indicate the chronological order of a series of events, as well as tasks that require the test taker to select a specified number of correct answers. Both formats were easy to score and provided useful information about the test taker's reading ability in that they discriminated well across proficiency levels. For constructed response, the unique answer format was most promising because it offers an alternative to constructed response items without increasing the amount of scoring time required. Unique answer items require the test taker to provide a short response consisting of a single number, word, or phrase and are constructed in such a way that there is only one plausible response, which greatly facilitates scoring the items.

Three of the reading/writing contexts also had associated writing tasks. These tasks were developed to assess test taker ability to generate a writing sample of a paragraph or more in length. Although other writing skills such as the ability to copy information or complete a form had been abstracted and identified by working group members as appropriate for placement, it was decided that it would be better to assess these skills at either the first or second tier of the placement process. Thus, the prototyping effort focused exclusively on the development of writing tasks that are communicative in nature and can be scored quickly and reliably.

---

[7]Constructed response item format requires the test taker to generate a response, while selected response item format requires the test taker to choose the correct answer(s) from a series of response options.

Pilot testing results indicated that all the writing tasks elicited ratable samples and did a credible job of discriminating across levels.[8] In addition, raters were able to score the tasks quickly and easily using holistic rubrics developed for both Level A and Level B tasks. (See Butler et al., 1996, for the Level A and Level B rubrics as well as a discussion of scoring procedures and rater training protocol.) Because all four writing tasks showed promise as models for placement, they were included in the specifications as prototypes along with sample responses to demonstrate application of the rubric.

Listening items were pilot tested in March 1995. Of the eight contexts pretested, seven were retained for pilot testing.[9] Table 3 presents the listening contexts pilot tested at Level A (BL-IH) and Level B (IL-AH) and indicates the number of items associated with each context. Level A listening contexts were pilot tested with 581 students from beginning low through intermediate high across six different agencies. Level B listening contexts were pilot tested with 410 students from intermediate low through advanced high across five agencies. Table 4 provides the number of students in the pilot administration of Level A and Level B listening contexts by proficiency level.

A variety of items using both constructed and selected response formats were developed to assess the following listening skills: extract specific information, draw meaning, extract global information, and make inferences. Many of the item formats that were found to be promising for assessing reading were used in developing listening items as well. An attempt was made to limit the amount of reading and writing required at Level A by developing predominantly picture-based items using selected-response formats. With Level B items, some reading and writing was required, but the language to be interpreted or produced was always at a lower level than the target level of the item.

---

[8]It should be noted, however, that few samples were found to match the advanced high writing descriptors, which may be due to the fact that writing had not previously been emphasized in the adult ESL curriculum.

[9]Some contexts were included in the pilot testing as a warm-up only. The use of warm-up items was recommended by the working group and proved to beneficial in orienting the test takers to the listening modality.

Table 3
Summary of listening contexts pilot tested in March 1995 by test level

Level	Listening Contexts	Number of Associated Listening Items
A*	Short Dialogues (brief conversations)	3
A	Short Monologues (descriptive sentences)	4
A	Extended Dialogue (conversation between doctor and patient appropriate for Level A)	8
A & B†	Short Monologues (brief recorded messages)	6
A & B†	Medium Monologue (long telephone message)	5
B	Extended Dialogue (conversation between doctor and patient appropriate for Level B)	10
B	Extended Monologue (news report)	11

*Intended as a warm-up for Level A

†Intended as a warm-up for Level B

Table 4
Number of students in the pilot administration of listening contexts by test level and Model Standards proficiency level

Test Level	Model Standards Proficiency Level						Total
	BL	BH	IL	IH	AL	AH	
A	154	140	151	136			581
B			91	124	127	68	410

Pilot testing results for listening indicated that the items generally did a good job of discriminating across levels, particularly the items intended to assess the ability to extract global information. Several items were developed to assess the ability to extract specific information from a listening text; some were targeted at Level A and others at Level B. Promising item formats for assessing this skill include matching and fill in the blank, for Level A, and unique answer for Level B. While the unique answer format showed promise, it also illustrated the complexity of assessing listening in an open-ended format because at times it was unclear whether test takers had understood the information retrieved or

had simply transcribed it. For these items, it was necessary to systematically review the responses in order to generate an appropriate scoring protocol. Although additional time was required to produce the protocol, it was time well spent in that these items provided important information about test taker listening ability at the upper levels.

A special pilot administration that involved the same students taking a subset of the reading, writing, and listening items was conducted in May 1995 to obtain preliminary information about student performance across skill areas. Because reading and writing items were developed and pilot tested separately from the listening items, this offered the first opportunity to examine same-student performance in all three skill areas. Pilot testing results from this administration helped address whole-test construction issues related to the differential performance of students across skill areas and the impact of such performance on the placement process. Appendix D provides the pilot testing results of this administration and discusses their implications.

An important component of both pre- and pilot testing was the collection of qualitative data in the form of observations, focus groups, and questionnaires. Observations of all pretest administrations were conducted to ascertain (a) the comprehensibility and familiarity of item formats, (b) the clarity of directions, and (c) the adequacy of the amount of time provided for students to complete the tasks. Student questions regarding item formats, directions, and unfamiliar vocabulary were recorded and used to inform revisions. In addition, project staff discussed the texts and items with participating students both on a whole-class level and, when possible, in small focus groups. Teacher feedback was obtained through individual discussions.

At the pilot testing stage, feedback was collected from teachers and students through the use of questionnaires (see Appendix C for a sample questionnaire). Students and teachers were asked questions regarding appeal and relative difficulty of texts and associated items.[10] In general, students across proficiency levels reacted positively. They felt the tasks were practical and provided them with a good opportunity to practice their

[10]Students did not individually complete questionnaires. Instead, the teachers asked the whole class a set of standardized questions and summarized student responses on an appropriate form.

English. In terms of difficulty, student responses varied across task and proficiency level, with lower-level students generally reporting greater difficulty in comprehending and completing tasks targeted at the upper levels and upper-level students generally reporting great ease in completing tasks targeted at the lower levels. This information provided initial evidence that many of the tasks were appropriate for the targeted levels, and when student reactions did not follow this pattern, project staff were alerted to potential problems.

In general, teachers also reacted favorably, but were often concerned that the tasks were too difficult for their students. However, pilot testing results indicated that the tasks were manageable for most of the population. This was particularly true with regard to writing: Many teachers feared that the writing tasks were too challenging, yet the majority of students produced ratable samples and responded positively to the tasks. The information gained from teachers and students at both stages of the development process was critical in assuring the overall quality of the prototypes.

The most promising items from the pilot testing were identified as prototypes and included in the specifications. The criteria for determining which items became prototypes were based on content considerations and statistical performance (i.e., the overall difficulty of the item for the sample and how well it discriminates between levels) and are presented in Butler et al. (1996).

The prototyping effort described above informed the test development plan in two important ways. First, it allowed for a variety of item types to be tried out, which provided insight into what was most effective for the target population and served as a catalyst for clarification of the specifications. The prototyping effort also provided project staff with an opportunity to try out an item development process which could serve as a model for future test developers. This information is included in the guidelines for test development summarized below.

Guidelines for Test Development

The guidelines presented in the test development plan indicate how the text and item specifications for reading, writing, and listening are to be used and provide direction for both item and whole-test development. The

guidelines are an outgrowth of the prototype development process and are strongly recommended to ensure the best use of the test development plan. While the complete set of guidelines is presented in Butler et al. (1996), the key points are summarized below.

The first step in the test development process is to assemble a core test development team including experts in language testing, ESL instruction, and psychometrics. The core team is responsible for overseeing the test development process and will work closely with an advisory committee consisting of one or more language testing experts and several representatives from agencies in California who will administer the tests once they are available. Language testing experts will provide input on further development of the specifications, on the plan for whole-test construction, and on issues of reliability and validity. Continued input from agency representatives will help verify the content and language appropriateness of texts and associated items and determine what is feasible operationally given the realities of individual agency situations.

Once the core test development team and the advisory committee are constituted, the text and item specifications for reading, writing, and listening should be carefully reviewed and completed where necessary. When a complete set of specifications is available, potential item writers can be identified and trained. Item writers should be familiar with the adult ESL population and with the Model Standards. Background in these two areas will facilitate the selection of appropriate material for text and item development.

The recommended item development process parallels the prototype development process described above. As in the prototyping effort, extensive review of texts and items at each stage of the development process is strongly recommended to assure the quality of the items produced and their match to the Model Standards. Similarly, multiple tryouts of texts and items are highly recommended to identify weaknesses in wording, directions, or formats that may impact the appropriateness of the items and their ability to discriminate across levels.

Once texts and item sets have been developed, a plan for whole-test construction must be drafted. The plan must address sampling issues and operational constraints, and incorporate the following whole-test decisions which were made in conjunction with the working group.

1. Both Level A and Level B instruments will assess test taker ability in three skill areas—reading, writing, and listening—with reading and writing in one section and listening in another.
2. Placement decisions will be based on performance in all three skill areas, although tests will be constructed in such a way that separate scores can be reported for each area.
3. Sampling of skills within a given skill area will be based on recommendations found in the item specifications for reading, writing, and listening.[11]
4. There will be a variety of item formats (both constructed and selected response); however, an attempt will be made to limit the number of different formats within a given section of the test.
5. Ease of scoring will remain a primary consideration in assembling whole tests. Example items should be included as needed as well as warm-up items at the beginning of the listening section to orient the test taker to the aural modality. Neither example items nor warm-up items should be scored.

At least two forms of both Level A and Level B tests should be assembled and all forms should be pilot tested to determine how each instrument is performing as a whole test. Timing adjustments can be made if necessary to assure that examinees have ample time to complete the test. Acceptable levels of reliability must be established and initial cutoff ranges for placement estimated. An operational test must be monitored regularly to determine if cutoff ranges are allowing for effective placement decisions vis-à-vis course content. Over time it should be possible to adjust cutoffs so that students are being placed into classes appropriate for their language ability.

Finally, steps must be taken to establish the validity of instruments produced from the specifications. It is important to stress that establishing validity is an ongoing process that begins in the initial stages and continues throughout the development of operational instruments. It will take the combined efforts of the core team and the advisory committee to assure the validity of the instruments to be produced.

---

[11]The recommendations found in the listening item specifications are based on preliminary discussions with the working group and will need to be systematically reviewed by the advisory committee.

Conclusion

The process that led to the creation of the specifications and the test development plan formally established the link between the Model Standards and the operational instruments to be produced from the specifications. Though the specifications in their current form are the result of a systematic development process, they must continue to evolve with use as item writers provide feedback regarding their effectiveness. Operational instruments produced from the specifications will help achieve the original goal of this project by providing additional options for a menu of tests appropriate for use with the Model Standards.

References

Bachman, L. (1990). *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Butler, F. A., Weigle, S. C., & Sato, E. Y. (1993). *California Department of Education Adult ESL Assessment Project. Final Report: Year 1.* Los Angeles: University of California, Center for the Study of Evaluation.

Butler, F. A., Weigle, S. C., Kahn, A. B., & Sato, E. Y. (1996). *California Department of Education Adult ESL Assessment Project. Test Development Plan with Specifications for Placement Instruments Anchored to the Model Standards.* Los Angeles: University of California, Center for the Study of Evaluation.

California Department of Education. (1992). *English-as-a-Second-Language Model Standards for Adult Education Programs.* Sacramento, CA: Author.

Kahn, A. B. (forthcoming). *The effects of task type on writing assessment: An adult ESL placement example.* Unpublished manuscript, University of California, Los Angeles.

Kahn, A. B., Butler, F. A., Weigle, S. C., & Sato, E. Y. (1994). *California Department of Education Adult ESL Assessment Project. Adult ESL Placement Procedures in California: A Summary of Survey Results.* Los Angeles: University of California, Center for the Study of Evaluation.

Wall, D., Clapham, C., & Alderson, J.C. (1994). Evaluating a placement test. *Language Testing, 11* (3), 321-344.

Weigle, S. C., Kahn, A. B., Butler, F. A., & Sato, E. Y. (1994). *California Department of Education Adult English-as-a-Second-Language Assessment Project. Final Report: Year 2.* Los Angeles: University of California, Center for the Study of Evaluation.

Weigle, S. C. (1995). *California Department of Education Adult English-as-a-Second-Language Assessment Project. Initial Cutoff Ranges for the BEST and NYS Place Test for Placement into Model Standards Proficiency Levels.* Los Angeles: University of California, Center for the Study of Evaluation.

Appendix A

Adult ESL Assessment Working Group Members

1994–1995

Adult ESL Assessment Working Group

Agency	Member
ABC Adult School, Cerritos	Jean Rose
Career Resources Development Center, San Francisco	Chris Shaw
City College of San Francisco	Nadia Scholnick
Fremont School for Adults, Sacramento	Mary White
Hayward Adult School	Joyce Clapp
Los Angeles Unified School District Division of Adult and Career Education	Barbara Martinez
Merced Adult School	Debbie Glass
Mt. Diablo Adult Education	Jacques LaCour
Oxnard Adult School	Judy Hanlon
San Diego Community College District Continuing Education Centers	Gretchen Bitterlin[a] Leann Howard[a]
Santa Clara Unified School District Educational Options	Bet Messmer
Torrance Adult School	Bertie Wood
Watsonville Adult School	Claudia Grossi

[a]For this agency two representatives shared responsibility.

Appendix B

General Content Grids:
Reading, Writing, and Listening

General Content Grids: Reading, Writing, and Listening

The General Content Grids for Reading, Writing and Listening are intended to serve as the content base for any type of assessment developed for use with the Model Standards—placement, diagnostic, progress, or exit. Six grids were developed: reading/writing text characteristics, reading skills, writing skills, listening text characteristics, listening skills, and language functions and forms.

For each grid, the appropriateness of text types, skills, or language functions and forms are indicated at the six decision points across levels as described in the Model Standards. A decision point is the boundary between two proficiency levels (e.g., between BL and BH). The appropriateness of texts and skills for a given decision point is determined operationally as follows. A text characteristic (e.g., length, topic) is considered appropriate at a decision point if it is mentioned explicitly in the Model Standards as part of the course content at the lower of the two levels comprising the decision point. In this case, a black box (■) is placed at the decision point for that text characteristic. Similarly, a black box placed at a decision point for a given skill means that the skill is an explicit part of course content at the lower of the two levels, indicating that students at the lower level will not have mastered the skill while students at the upper level will have done so. Functions and forms use a similar notation, with black boxes indicating that a function or form is specified in the Model Standards as course content for the lower level at a decision point.

Because not all areas are completely delineated in the Model Standards at all levels, white boxes (❑) are used to indicate that the text characteristic, skill, function, or form may be appropriate for test construction at that decision point, even though it is not explicitly mentioned in the Model Standards at the lower level.[12] Finally, it should be noted that the level of text and item difficulty also depends on factors such as familiarity of topic, visual aids, and complexity of content, vocabulary, and syntax, as specified in Model Standards.

---

[12]The placement of white boxes was verified by working group members as part of an initial validation process.

Reading/Writing Text Characteristics

	DECISION POINT	BL/ BH	BH/ IL	IL/ IH	IH/ AL	AL/ AH	AH/ +
Length							
	word or phrase	■	■	■			
	sentence	■	■	■			
	paragraph		■	■	■	❑	❑
	passage		■	■	■	■	■
Topic/Type							
general[a]	lists, menus, directories, indices	■	■	❑			
	calendars, schedules	■	■	■			
	signs, labels	■	■	■	❑		
	advertisements	■	■	■	❑	❑	❑
	forms	■	■	■	■	■	■
	tables		❑	❑	❑	■	❑
	public information notices		❑	❑	■	❑	
	notes, messages, letters		■	■	■	■	❑
	newspaper/magazine articles		❑	■	■	■	■
	consumer materials			❑	■	■	■
	prose fiction (short stories, fables)			❑	■	■	■
vocational[b]	advertisements	❑	■	■			
	paychecks		❑	■	■		
	labels		❑	❑	❑	❑	❑
	forms		■	■	■	■	❑
	letters, memos, reports, logs			❑	■	❑	■
	technical materials			❑	❑	■	■
	resumes			❑	❑	❑	■
academic[c]	newspaper/journal articles			❑	■	■	❑
	biographies			■	❑	❑	❑
	tables, charts, graphs			❑	❑	■	❑
	forms			❑	❑	❑	❑
	technical documents				❑	❑	■
	textbooks				■	■	■
	literary texts				■	■	■

[a]General topics may include shopping, banking, housing, health, transportation, current events, community resources, and other personal matters.

[b]Vocational topics may include employment, customer relations, benefits, wages, and safety.

[c]Academic topics may include literature, science, history, government, commerce, and intercultural issues.

BEST COPY AVAILABLE

Reading Skills

	DECISION POINT	BL/ BH	BH/ IL	IL/ IH	IH/ AL	AL/ AH	AH/ +
Skills							
locate	non-alphabetic information	■	■	■			
	alphabetic information	■	■	■			
draw meaning	from a proposition	■	■	■	■	■	■
	from a series of propositions		■	■	■	■	■
extract & combine information	from different sections in a text		❑	❑	■	❑	❑
	from different texts		❑	❑	❑	❑	❑
interpret relationships	cause/effect				■	❑	❑
	compare/contrast				■	❑	❑
	generalization/example				■	❑	❑
	main idea/supporting details			❑	■	■	■
	sequence of events		■	■	■	■	
analyze	make inferences (recognize point of view, draw conclusions)			❑	❑	■	■
	distinguish fact from opinion				❑	❑	❑
	identify rhetorical structure					❑	■

Writing Skills

	DECISION POINT	BL/ BH	BH/ IL	IL/ IH	IH/ AL	AL/ AH	AH/ +
Length of Expected Response							
	word or phrase	■	■				
	sentence	■	■				
	series of related sentences (paragraph)		■	■	■		
	series of short paragraphs				■	■	■
Skills							
copy	familiar written material (e.g., lists, recipes, directions, stories)	■	■				
transcribe/ take notes	familiar material transmitted orally (e.g., recipes, messages, directions)	■	■	■	❑		
	simple notes from short lectures, public announcements, or interviews				❑	■	❑
	notes from academic lectures					❑	■
complete	short, simple forms requesting routine information (e.g., name, address, phone)	■	❑				
	simple forms requesting detailed biographical or personal information		❑	■	■		
	specialized forms requesting specific, detailed information				❑	■	❑
generate	notes, messages		■	■	❑		
	letters, memos			❑	■	■	■
	prose	■	■	■	■	■	■
	narration	❑	❑	❑	■	■	❑
	description	❑	❑	❑	❑	■	❑
	exposition					❑	■
	simple outlines					❑	■
	short summaries					❑	■
use	rhetorical techniques		❑	❑	■	■	■
	chronological order		❑	❑	■	■	❑
	comparison/contrast					❑	■
	cause/effect					❑	■
	generalization/example					❑	■

Listening Text Characteristics

	DECISION POINT	BL/ BH	BH/ IL	IL/ IH	IH/ AL	AL/ AH	AH/ +
Topic	general[a]	■	■	■	■	■	■
	vocational[b]	■	■	■	■	■	■
	academic[c]			■	■	■	■
LISTENER CAN INTERACT WITH THE SPEAKER							
Modality	visual (e.g., face-to-face conversations)	■	■	■	■	■	■
	nonvisual (e.g., phone conversations)		■	■	❑	❑	❑
Length	brief	❑	❑	❑			
	extended			❑	❑	❑	❑
LISTENER CANNOT INTERACT WITH THE SPEAKER							
Modality	visual (e.g., TV, movies, lectures)	■	■	■	■	■	■
	nonvisual (e.g., radio, recorded phone information, public announcements)	■	■	■	■	■	■
Two or more speakers							
Length	short (fewer than 50 words)	❑	❑	❑			
	medium (50-100 words)		❑	❑	❑	❑	
	long (101-250 words)			❑	❑	❑	❑
	extended (251-350 words)			❑	❑	❑	❑
One speaker							
Type	recorded phone information	❑	■	❑	❑		
	public announcements	■	■	❑	❑	❑	❑
	stories			❑	■	❑	❑
	lectures, speeches				❑	■	■
	broadcasts				❑	■	❑
Length	word or phrase	■	■				
	single sentence	■	■	■			
	short passage (1 paragraph)	❑	■	■	■	❑	❑
	long passage (2 or more paragraphs)			❑	■	■	■

[a]General topics may include shopping, banking, housing, health, transportation, current events, community resources, and other personal matters.

[b]Vocational topics may include employment, customer relations, benefits, wages, and safety.

[c]Academic topics may include literature, science, history, government, commerce, and intercultural issues.

Listening Skills

DECISION POINT	BL/ BH	BH/ IL	IL/ IH	IH/ AL	AL/ AH	AH/ +
extract specific information (single word or phrase)	■	■	❑	❑	❑	❑
draw meaning	■	■	■	■	■	■
extract global information	■	■	■	■	■	■
make inferences (e.g., place, mood)		❑	❑	■	❑	❑

Language Functions and Forms

	DECISION POINT	BL/ BH	BH/ IL	IL/ IH	IH/ AL	AL/ AH	AH/ +
Functions							
factual	tell/describe/identify/explain/illustrate	■	■	■	■	■	■
	express modality (necessity, obligation, certainty, ability, possibility)		■	■	■	■	■
	compare/contrast, conclude, infer, evaluate, analyze				■	■	■
social	basic social functions (introduce, greet, take leave, compliment, apologize, etc.)	■	■	■	■	■	■
	express emotion (state of being, desire, worry, hope, regret, satisfaction, etc.)	■	■	■	■	■	■
suasive	request, direct, invite	■	■				
	suggest, advise, recommend, persuade			■	■	■	■
	solve problems, predict consequences				■	■	
Sentence Types							
simple	affirmative & negative statements	■					
	yes/no, *or*, & *wh-* questions & answers	■					
	commands	❑					
	direct speech		❑	■			
	exclamatory sentences		❑	■			
	tag questions			❑	❑	❑	■
compound	and, but, or, and...too, and.... either	■	■				
complex	adverb clauses (time, reason, concession)		❑	■	❑	■	
	adjective clauses			❑	■		
	indirect speech & embedded questions			❑	■		
	sentences with conjunctive adverbs				❑	■	
	noun clauses					❑	■
	present subjunctive					❑	■
Verb Forms							
simple	present, past, future	■	■				
	infinitives	❑	■	❑			
	gerunds		❑	■			
modals	can, have to, could, should, must, may, would, might, used to	■	■	■	❑	❑	
	past forms (should have, could have, etc.)				❑	■	
complex	continuous (present, past, future)	■	❑	■	❑	❑	
	perfect (present, past, future)		❑	■	■	❑	■
	perfect continuous (present, past, future)		❑	■	■	❑	■
	conditional (future, contrary to fact, past, continuous)		❑	■	■	■	■
	passive (simple present, past, future)			❑	■	■	
	causative				❑	■	

BEST COPY AVAILABLE

Appendix C

Sample Teacher/Student Questionnaire

This appendix presents a sample questionnaire used in the October 1994 pilot testing of reading and writing items and reflects the kind of information obtained from teachers and students in individual interviews and in small focus groups.

Teacher/Student Questionnaire

Please answer the questions below. After administration of the exercise, ask your students the questions on the back of this form and record their comments.

Name of Agency: _____

Teacher's Name: _____

Class Level: _____

For Teachers

Do the tasks in the reading and writing exercise booklet reflect skills that are taught at your class level?

Are there any items that you particularly liked or disliked? Please explain.

Did the students appear to understand all instructions? Which items, if any, seemed especially problematic for students?

How much time was needed for most students to finish the exercise? What was the range of time spans needed to finish?

Other comments

For Students

Did you like this reading and writing exercise? Why or why not?

Did you understand what you were supposed to do? Were the examples helpful?

Which questions did you like the best? Why?

Which questions were the most difficult for you? Why?

Do you think this is a good way to test your English reading and writing? Why or why not?

Other comments

Appendix D

Special Pilot Administration of Reading, Writing, and Listening Items

Special Pilot Administration of Reading, Writing, and Listening Items

In May 1995 Level A and B listening, reading, and writing items were administered to the same students at one site to obtain preliminary information about how students at different levels perform across the skill areas. Table D1 presents the number of items administered in each skill area for Levels A and B, and Table D2 shows the number of students at each proficiency level who took part in the test administration.[13]

Table D1
Number of items in May 1995 pilot administration by test level and skill area

	Listening	Reading	Writing
Level A	26	16	1
Level B	28	11	1

Table D2
Number of students in May 1995 pilot administration of reading, writing, and listening items by test level and Model Standards proficiency level

	BH	IL	IH	AL	Visa 6'	Visa 7'	Total
Level A	54	43	43				140
Level B		38	26	33	21	18	136

'Students more proficient than AH

The descriptive statistics for the Level A and Level B administrations are found in Tables D3 and D4, respectively. The tables show that scores on the Level A reading and listening items increase steadily from beginning high through intermediate high, suggesting that these items are useful for discriminating among these levels. The writing scores for Level A show an increase from beginning high to intermediate low, but not from intermediate low to intermediate high. Because of the limited sample, it is impossible to ascertain whether the task itself does not discriminate

[13]Reading, writing, and listening exercises were also administered to visa students (adult ESL students who are more proficient than AH and thus resemble exit-level students) to gauge the appropriateness of the tasks for the population.

between the two levels or whether the writing scores reflect a lack of emphasis on writing in the curriculum.

For Level B, the scores in all three skill areas show an increase from intermediate high to advanced low, and from advanced low to the two visa levels, but not from intermediate low to intermediate high. Again, the limited sample is not sufficient to ascertain whether the lack of discrimination between intermediate low and intermediate high is a function of the items themselves or of the placement of students with similar abilities in both levels.

Table D3

Descriptive statistics for May 1995 pilot administration of Level A items by Model Standards proficiency level and skill area

		Listening[a]			Reading[b]			Writing[c]		
	N	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.	Range
BH	54	17.94	4.32	5-24	8.11	3.21	1-14	1.61	1.20	0-4
IL	43	19.67	3.70	9-24	10.51	2.96	5-15	2.47	1.20	0-4
IH	43	21.81	2.36	16-25	11.81	2.31	6-16	2.47	.99	0-4
Total	140	19.66	3.94	5-25	9.99	3.27	1-16	2.14	1.21	0-4

[a]Total number of items = 26
[b]Total number of items = 16
[c]Total number of items = 1; score range = 0–4

Table D4

Descriptive statistics for May 1995 pilot administration of Level B items by Model Standards proficiency level and skill area

		Listening[a]			Reading[b]			Writing[c]		
	N	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.	Range
IL	38	14.61	3.51	6-21	5.58	2.65	2-11	3.01	1.33	0-5
IH	26	13.04	4.00	5-21	5.38	2.76	1-10	3.08	1.00	0-4
AL	33	18.30	3.15	12-24	7.06	2.01	2-11	3.76	.64	3-5.5
Visa 6[d]	21	24.90	2.56	21-28	10.00	1.10	7-11	5.05	.86	3-6
Visa 7[d]	18	26.17	1.42	23-28	10.17	1.04	7-11	5.00	.86	3-6
Total	136	18.32	5.84	5-28	7.19	2.90	1-11	3.78	1.30	0-6

[a]Total number of items = 28
[b]Total number of items = 11
[c]Total number of items = 1; score range = 3–6
[d]Students more proficient than AH

Correlations among the three skill areas for Level A and Level B are found in Tables D5 and D6, respectively. For Level B, students at the two visa levels have been excluded from the correlations because they do not represent the population for whom the items are intended. As the tables show, the three skill area item sets are significantly correlated with each other at both levels, with listening and reading correlated more strongly than either correlates with writing. The higher correlations between listening and reading may be due to a test method effect (multiple-choice and unique answer vs. composition) (Bachman, 1990), or may simply reflect the fact that listening and reading tend to be addressed more than writing both in the current placement process and in the curriculum, and thus may develop at a similar rate. In any case, the correlations are of an appropriate magnitude for placement purposes: neither so low that the items in the different skills seem to be measuring completely unrelated abilities, nor so high that they are providing redundant information (Wall, Clapham, & Alderson, 1994). In fact, the lower correlations between writing and the other two skill areas argue for including writing in the placement process since the writing scores tend to give somewhat different information about student abilities than do reading or listening.

Table D5

Correlations among skill areas for May 1995 pilot administration of Level A reading, writing, and listening items

	Listening	Reading	Writing
Listening	1.00		
Reading	.73**	1.00	
Writing	.66**	.61**	1.00

**$p \leq .01$

Table D6

Correlations among skill areas for May 1995 pilot administration of Level B reading, writing, and listening items†

	Listening	Reading	Writing
Listening	1.00		
Reading	.61**	1.00	
Writing	.49**	.49**	1.00

**$p \leq .01$

†IL through AL only